DOCUMENT RESUME

ED 120 260 TM 005 223

AUTHOR Pfau, Richard H.
TITLE The Cross-National Comparison of Classroom Behaviors.
PUB DATE [Feb 76]
NOTE 32p.; Paper presented at the Annual Meeting of the Comparative and International Education Society (20th, Toronto, Canada, February 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage
DESCRIPTORS Behavior; Classification; *Classroom Observation Techniques; Classrooms; *Comparative Education; *Cross Cultural Studies; Validity

ABSTRACT

This paper indicates why the use of "Category Systems" provides a potential for making precise and valid cross-cultural comparisons of classroom behaviors not equaled by other techniques. The paper is presented to highlight a technique which seems especially suitable for use by comparative educators and other social scientists who wish to make precise comparisons of classroom behaviors, to indicate why this technique is apparently more suitable than others, and to stimulate discussion leading to an increase in the number and sophistication of cross-cultural studies of classroom behaviors. Although this presentation focuses upon the cross-national comparision of classroom behaviors, the technique discussed is equally applicable to comparisons of teaching associated with different ethnic groups within a single country, and to ohter types of comparisons as well. Examples are given of several category systems and data resulting from their use in the United States and Nepal. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED120260

THE CROSS-NATIONAL COMPARISON OF CLASSROOM BEHAVIORS

Richard H. Pfau
International and Development Education Program
University of Pittsburgh
Pittsburgh, Pa. 15260
U. S. A.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Abstract

This paper indicates why the use of "category systems" provides a potential for making precise and valid cross-cultural comparisons of classroom behaviors not equaled by other techniques. Examples are also given of several category systems and data resulting from their use in the U.S.A. and in Nepal.

Paper presented at the XXth Annual Convention
of the Comparative and International Education Society
Toronto, Canada, 24-27 February 1976

IM005 223

Introduction

Comparative education as a discipline is concerned with the study of cross-national and cross-cultural variability in the educational domain (Husen, 1967, Vol.I, p.19). However, at the present time, comparative educators have only vague ideas about what occurs in the classrooms of most countries, and how classroom behaviors vary from one social system to another.

This paper is presented (1) to highlight a technique which seems especially suitable for use by comparative educators and other social scientists who wish to make precise comparisons of classroom behaviors, (2) to indicate why this technique is apparently more suitable than others, and (3) to stimulate discussion leading to an increase in the number and sophistication of cross-cultural studies of classroom behaviors. Although this presentation focuses upon the cross-national comparison of classroom behaviors, the technique discussed is equally applicable to comparisons of teaching associated with different ethnic groups within a single country, and to other types of comparisons as well.

Reasons for making such comparisons, across cultures and nations, include (1) determination of the generality of classroom related theory (Comber & Keeves, 1973, p.299; Warwick & Osherson, 1973, p.8), (2) generation of theory (Medley, 1967), (3) identification of variability across cultures to obtain otherwise unavailable experimental treatments (Brislin, Lonner, & Thorndike, 1973, p.7; Husen, 1967, Vol.I, pp.21-22). and (4) provision of information about classroom occurrences to educational planners and evaluators.

Techniques for Measuring and Comparing Behaviors

Social scientists wishing to compare classroom behaviors require data about those behaviors in order to make comparisons. This data can be gathered by using one or more of a variety of techniques.

Often, educational researchers use relatively indirect techniques, such as interviews and questionnaires, to gather information about classroom activities. Such techniques require that teachers or students report about what occurs in their classes. The resulting answers are then used as a basis for describing classroom occurrences and for comparing occurrences between groups of classrooms.

The International Association for the Evaluation of Educational Achievement (IEA) studies of mathematics, science, and other subjects used indirect techniques in an attempt to determine what happens in the classrooms of the countries studied. Science teachers, for example, were sent a Science Teacher Questionnaire which contained questions such as:

> Indicate how often you give your Science students opportunities for planning and carrying out limited scientific investigations on their o[illegible]. (Comber & Keeves, 1973, pp.275-276)

Four categories of response were used for such questions: "never," "seldom," "occasionally," and "frequently."

Similarly, in a major comparative study of teacher roles,

> respondents were asked to indicate for each item the amount of emphasis given it in their own classroom teaching. Five response categories were provided: "A great deal of emphasis", "Strong emphasis," "Moderate emphasis," "Slight emphasis" and "Little or no emphasis." (R.S. Adams, 1970, p.51)

The responses to such questions were used to infer the

magnitudes of classroom behaviors.

Major difficulties are faced by researchers using such techniques cross-culturally, and serious questions can be raised about the validity of inferences drawn from the use of such measures. For example, does the term "frequently" mean the same thing to a British and an Indian teacher, and thus are the responses given by these teachers actually comparable? Similarly, does the term "moderate emphasis" mean the same thing to persons of differing cultural backgrounds, and does it elicit responses implying similar meanings?

The difficulties encountered by using such indirect techniques of measurement can be appreciated by looking at the writings of researchers who have used these techniques. These researchers reach conclusions such as "the extent to which these reports are veridical (that is, they reflect actual teaching practices) is an open question" (R.S. Adams, 1970, p.53); "response bias was operating -- in other words, some countries were being freer with their willingness to emphasize (anything) than were others" (p.52); "what our respondants reported then, might represent their values rather than their practices" (p.58); and

> the teachers, though responding to the same word cues may have placed different meaning on them. For example, "free communication" may be semantically quite different in the United States and Australia. Again, teachers may be poor perceivers of their own performance. Thus their reports do not reflect the reality of their teaching." (R.S. Adams, 1970, p.58)

To further appreciate the number and magnitude of potential problems involved in making valid comparisons using indirect techniques, one need only look at the many chapters relating

to the use of questionnaires and interviews contained in cross-cultural and comparative research methods books.[1]

If one's purpose is to describe and compare classroom behaviors, a potentially more valid technique is to have trained persons directly observe such behaviors, record their observations, and then use those records as a basis for making comparisons. Direct classroom observations, however, can be recorded and reported in a number of different ways.

One method, often used by educators and anthropologists, is to observe classroom teaching and then write descriptive accounts of the teaching observed. An example of a description resulting from such an approach is the following:

> Teaching techniques seldom varied; the teacher might read aloud from a textbook while some children took a few notes, or the teacher or a child would chant a standard question and the class would respond by chanting a memorized answer, or the teacher would make a statement and the class would repeat it in unison. (Reed, 1968, p.135)

Such descriptions, however, do not provide a very precise basis for comparing patterns of teaching from one country to another, or indeed from one school to another, since they furnish only gross indications of the actual extent to which specific behavior patterns occur. For example, the above description gives no indication of how much time teachers spent reading aloud from textbooks -- was it 50% of a class period, 20%, 1%?

Furthermore, such descriptions suffer from the problem that the descriptive words employed lack a normative base for making comparisons. That is, they lack an explicit language of comparison. Thus, as Deutscher (1973) has pointed out, the standard for the same objects or behaviors

may well vary from culture to culture, from nation to nation, or for that matter, within any given social unit -- between classes, age groups, sexes, and so on. "What is 'cold' soup for an adult may be too 'hot' to give to a child" (Deutscher, 1973, p.174). Similarly, what is "seldom" to one person may appear "frequent" to another.

It is apparent that more systematic observation, recording, and reporting procedures are required if precise descriptions and comparisons of classroom behaviors are to be made.[2]

Systematic classroom observation instruments which now exist are of generally two types, these being "rating systems" and "category systems."[3] In the past, rating systems were distinguished from category systems mostly by the amount of inference inherent in their use and in the interpretation of their results. As Rosenshine (1970) has pointed out:

> Category systems are classified as low-inference measures...because the items focus upon specific, denotable, relatively objective behaviors such as "teacher repetition of student ideas," or "teacher asks evaluative question," and because these events are recorded as frequency counts. Rating systems are classified as high-inference measures because they lack such specificity. Items on rating instruments such as "clarity of presentation," "enthusiasm," or "helpful toward students" require that an observer infer these constructs from a series of events. In addition, an observer must infer the frequency of such behaviors in order to record whether it occurred "consistently," "sometimes," or "never," or whatever set of gradations are used in the scale of an observation instrument. (p.281)

Although rating systems are no longer limited to high-inference items, and high-inference items have been used in some category systems (Rosenshine & Furst, 1973, p.133), the amount of inference inherent in the use of these instruments still serves as a useful distinguishing feature

in most cases.

To further help in distinguishing between category and rating systems, it can be kept in mind that observers using category systems keep a record of specific events each time they occur, or at very frequent intervals (e.g., every 3 to 5 seconds). This usage contrasts with observers using rating systems, who usually estimate the frequency of events and extent of attributes only once, at the end of an observation session (Rosenshine & Furst, 1973, p.132). Whereas rating systems typically call for high-inference judgements, requiring that an observer integrate whatever he has witnessed over one or more periods of observation and provide a record of general impressions, observers using category systems must make specific judgements while the activities of the classroom are in progress (Dunkin & Biddle, 1974, p.60).

Rating systems suffer from a number of inherent defects. As indicated before, "ratings" are estimates of the degree to which a person or thing possesses a given characteristic (Remmers, 1963, p.329). These estimates are made and recorded usually once, at the end of a period of observation. Medley and Mitzel (1963), however, point out that it is desirable that behaviors be recorded as soon after they occur as possible, for "it is known that many factors can affect memory and may seriously distort a record made in retrospect" (p.251). These distorting factors result in what Kerlinger calls "the intrinsic defect of rating scales," this being their proneness to constant or biased error (Kerlinger, 1964, p.516).[4]

When different recording biases occur, as they are likely to, in different ways by persons with different cultural backgrounds, the utility of rating systems for making cross-

cultural comparisons is seriously undermined. Add to these problems the difficulty of providing operational definitions of the high-inference concepts used in most rating systems, and it can be seen that studies which call for observers to use rating systems may result in judgements that are also unreliable, as well as biased.

I feel that category systems, rather than rating systems, are potentially more suitable for gathering data to be used for making cross-national comparisons. The low-inference procedures associated with the use of most category systems seem to lend themselves better to more standardized recording of information by observers of differing cultural backgrounds. This more standardized recording, in turn, can provide information which is more appropriate for the making of valid cross-national comparisons.

Thus, of the various techniques available for gathering data about classroom behaviors, category systems seem capable of providing more precise, reliable, and comparable data than other techniques which now exist.

Sample Category Systems and Comparative Data

So far, the discussion of category systems and their appropriateness for cross-national comparative purposes has been a bit abstract. Brief descriptions of two category systems and comparative data resulting from their use are given below, in order to provide a more concrete basis upon which the suitability of such systems for comparative research purposes can be considered.[5]

Table 1 contains a description of a category system known

Table 1

Activity Categories Instrument (Science)

1. LABORATORY EXPERIENCES: OPEN-ENDED	Students are presented a problem to be solved by experimentation. The procedure may or may not be given. They are required to make observations and analyze or interpret their findings.
2. LABORATORY EXPERIENCES: STRUCTURED	Students are presented a laboratory experiment with a structured procedure. They are not required to analyze or interpret their data. They are asked to make observations.
3. GROUP PROJECTS	One or more groups of students are working on a science project during the class period. Some may work individually. (Not written projects).
4. STUDENT DEMONSTRATIONS	A student or group of students demonstrate a science experiment or project which they have prepared. (Oral report on science project would be included).
5. LIBRARY RESEARCH, REPORTING, FIELD TRIPS	(a) A student or group of students give an oral report they have prepared based on reference material. (b) The class works with reference materials for purposes of writing or making reports. (c) Field trips.
6. STUDENT SPEAKING	The student contributes verbally by asking a question, answering a question, or simply volunteering information. Student writing on the blackboard is also included.
7. TEACHER QUESTIONING	Students are asked a question by the teacher. Silence following questions is also included.
8. WORKBOOK WORK	Students work in class on workbooks, homework, questions from text, art-type work, etc.
9. TEACHER DEMONSTRATIONS	The teacher presents materials by demonstration, models, charts, textbook pictures, film, TV, filmstrip, radio, record, etc.
10. LECTURE	The teacher reads aloud, writes on blackboard, expresses his views, gives directions, makes an assignment or asks rhetorical questions. Students are expected to listen. They may interrupt only when they do not understand. Student reading in the text is also included.
11. SILENCE OR GENERAL HAVOC	The class may be cleaning up, settling down, or doing nothing. In general, this category should be used sparingly.

Note. This category system was developed by Caldwell (1967, 1968) and modified slightly by Pfau et al. (in press).

as the Activity Categories Instrument (ACI), developed by Caldwell (1967, 1968). As can be seen, this instrument consists of 11 categories of behavior which describe types of activities occurring in science classes. An observer using this instrument makes a record of classroom activities with the aid of a stopwatch. Every five seconds the observer notes a single number associated with one of the ACI categories of behavior. The activity which occurs during each five-second interval determines which category number is recorded. Thus, if the teacher was asking a question, the observer records a 7, (teacher questioning). If a student response to the question occupied the next five-second interval, the number 6 (student speaking) would be recorded. In this way a series of numbers is recorded that indicates which of the 11 categories of behavior occurred during a period of observation, and the relative frequency of their occurrence.

Table 2 describes another category system known as the Flanders Interaction Analysis Categories (FIA).. This instrument describes 10 categories of behavi which refer mainly to types of student and teacher tal he FIAC can be used to describe the talk occurring formal classes of any subject matter area. An obs using this instrument records a category number representing the behavior observed about every three seconds. Observers are sufficiently trained to do this without the aid of watches. The numbers are recorded in sequence on an observation sheet. A recording of ...5,5,4,8,2,4,..., for example, would indicate that the teacher lectured, asked a question which was followed by a student response, praise or encouragement was then given by the teacher, and he asked another question.

Observers using category systems, such as these, are usually

Table 2

Flanders Interaction Analysis Categories (FIAC)*

Teacher Talk	Response	1. *Accepts feeling.* Accepts and clarifies an attitude or the feeling tone of a pupil in a nonthreatening manner. Feelings may be positive or negative. Predicting and recalling feelings are included. 2. *Praises or encourages.* Praises or encourages pupil action or behavior. Jokes that release tension, but not at the expense of another individual; ... saying "Um hm?" or "go on" are included. 3. *Accepts or uses ideas of pupils.* Clarifying, building, or developing ideas suggested by a pupil. Teacher extensions of pupil ideas are included but as the teacher brings more of his own ideas into play, shift to category five.
		4. *Asks questions.* Asking a question about content or procedure, based on teacher ideas, with the intent that a pupil will answer.
	Initiation	5. *Lecturing.* Giving facts or opinions about content or procedures; expressing *his own* ideas, giving *his own* explanation, or citing an authority other than a pupil. 6. *Giving directions.* Directions, commands, or orders to which a pupil is expected to comply. 7. *Criticizing or justifying authority.* Statements intended to change pupil behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.
Pupil Talk	Response	8. *Pupil-talk—response.* Talk by pupils in response to teacher. Teacher initiates the contact or solicits pupil statement or structures the situation. Freedom to express own ideas is limited.
	Initiation	9. *Pupil-talk—initiation.* Talk by pupils which they initiate. Expressing own ideas; initiating a new topic; freedom to develop opinions and a line of thought, like asking thoughtful questions; going beyond the existing structure.
Silence		10. *Silence or confusion.* Pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

*There is *no* scale implied by these numbers. Each number is classificatory; it designates a particular kind of communication event. To write these numbers down during observation is to enumerate, not to judge a position on a scale.

(Flanders, 1970, p.34)

trained until their recordings agree highly with those made by experts in the use of the instruments. This high agreement helps ensure that observations are made in a standardized way. Also, the fact that high agreement can be reached indicates that the process of assigning category numbers to describe observations is unambiguous.

Once recordings are made, they can be analyzed in a number of ways. For example, frequency counts can be made to determine how many times each category number was recorded during the period of observation. Based upon these, the percentage of time during which each category of behavior occurred can be estimated. If observations are made of a sample of classes, means or medians can be calculated and compared with those of other samples.

These standardized recording and analysis procedures permit us to make relatively precise statements about the occurrence of specific classroom behaviors. Table 3, for example, contains statistics descriptive of behaviors occurring within Nepal and the U.S.A. By looking at this table, we can state with some assurance that more student speaking occurred in the 5th grade science classes observed in the U.S.A. than in the Nepalese classes observed, more teacher questioning occurred in the U.S.A., more lecture (as defined by Caldwell) occurred in Nepal, the U.S. teachers made more use of audio-visual materials than did Nepalese teachers, and no laboratory experiences, group projects, library work or field trips occurred in Nepal, whereas such activities did occur in the U.S. classes observed.

Table 4 contains calculations indicating the percent occurrence of the 10 FIAC categories of behavior in Nepalese 5th grade and U.S. 6th grade social studies classes. In

Table 3

Mean ACI Calculations of Nepalese and U.S. Fifth Grade Science Classes

ACI Category	Nepal[a]	U.S.[b]
1. Laboratory experiences: open-ended	0.0 (0.0)	0.3 (1.2)
2. Laboratory experiences: structured	0.0 (0.0)	1.6 (3.7)
3. Group projects	0.0 (0.0)	1.3 (6.2)
4. Student demonstrations	0.0 (0.0)	1.8 (3.3)
5. Library research and field trips	0.0 (0.0)	3.8 (9.4)
6. Student speaking	6.8 (5.8)	18.5 (8.8)
7. Teacher questioning	6.7 (6.2)	17.1 (7.4)
8. Workbook work	1.9 (5.8)	4.4 (9.1)
9. Teacher demonstrations	0.3 (0.6)	5.0 (4.1)
10. Lecture	78.3 (12.2)	40.2 (11.0)
11. Silence or general havoc	5.9 (4.1)	3.9 (2.3)

Note. All numbers represent percent occurrences; numbers in parentheses indicate standard deviations.

[a]The Nepalese statistics shown are based upon observation of classes conducted during 1974 in a representative sample of 23 schools located in central Nepal.

[b]The U.S. statistics shown are based upon observation of classes taught by 30 teachers of the Syracuse, New York, area school system during 1966 (Caldwell, 1968).

Table 4

Mean FIAC Calculations of Nepalese and U.S. Fifth and Sixth Grade Social Studies Classes

Variable	Nepal[a]	U.S.[b]
FIAC category		
1. Accepts feelings	0.0 (0.1)	0.2 (0.2)
2. Praises or encourages	2.1 (2.5)	0.7 (0.7)
3. Accepts or uses ideas of students	0.1 (0.2)	7.6 (3.5)
4. Asks questions	6.2 (4.6)	9.2 (4.2)
5. Lecturing	72.4 (15.2)	27.2 (11.6)
6. Giving directions	1.7 (1.3)	3.6 (1.9)
7. Criticizing or justifying authority	0.6 (0.7)	2.3 (2.4)
8. Student talk - response	6.2 (5.0)	24.4 (14.8)
9. Student talk - initiation	1.9 (2.4)	13.8 (6.6)
10. Silence or confusion	8.7 (4.6)	11.0 (5.1)
Total talk		
Teacher talk	83.2 (8.7)	50.8 (13.7)
Student talk	8.1 (5.8)	38.2 (14.9)

Table 4 (continued)

Variable	Nepal	U.S.
S/(S+T) ratio	.090 (.067)	.427 (.159)
Indirect - direct indicators		
I/(I+D) ratio	.108 (.101)	.356 (.122)
i/(i+d) ratio	.418 (.257)	.578 (.184)
Extended indirect influence	0.1 (0.1)	2.0 (1.7)
Extended direct influence	0.2 (0.3)	2.4 (2.1)
Rapidity of behavior change		
Steady state ratio (10x10)	71.3 (15.7)	54.9 (6.9)
Pupil steady state ratio	14.1 (19.8)	58.5 (15.1)
Other calculations		
Cell 9-9	0.7 (1.5)	8.0 (5.4)
Pupil initiation ratio	21.3 (23.4)	39.0 (19.9)
Content cross ratio	87.9 (7.8)	49.0 (14.1)
Drill	4.7 (4.0)	6.0 (2.8)
Cell 3-3	0.0 (0.0)	1.7 (1.4)
Teacher question ratio (TQR)	8.8 (7.9)	27.1 (12.4)

Table 4 (continued)

Variable	Nepal	U.S.
Instantaneous teacher question ratio (TQR89)	22.6 (20.4)	42.2 (16.9)
Instantaneous teacher response ratio (TRR89)	75.1 (26.2)	82.4 (14.8)

Note. Numbers in parentheses indicate standard deviations.

[a]The Nepalese statistics shown are based upon observation of fifth grade classes conducted during 1974 in a representative sample of 21 schools located in central Nepal.

[b]The U.S. statistics shown are based upon observation of sixth grade classes taught by 30 teachers in the State of Michigan during 1964-65 (Flanders et al., 1969).

addition to percent descriptions of the 10 basic categories, a number of other more sophisticated descriptive indices are presented in this table, in order to indicate the variety of information which can be obtained by using one relatively simple category system. A person familiar with the meanings of these indices (see Appendix A for explanations) and associated statistics would interprete them to mean that the Nepalese teachers talked about ten times more than their students, whereas the U.S. teachers talked only slightly more than their students; Nepalese student talk tended to be quite short, usually less than three seconds in length, in contrast to the talk of U.S. students; a higher proportion of U.S. student talk was student initiated; Nepalese teachers put relatively more emphasis on lecturing and less on questioning students than U.S. teachers; U.S. teachers exhibited more extended direct behaviors (criticizing students, giving directions) and more extended indirect behaviors (praising and encouraging, accepting or using student ideas) than Nepalese teachers who rarely displayed such behaviors; and so on.

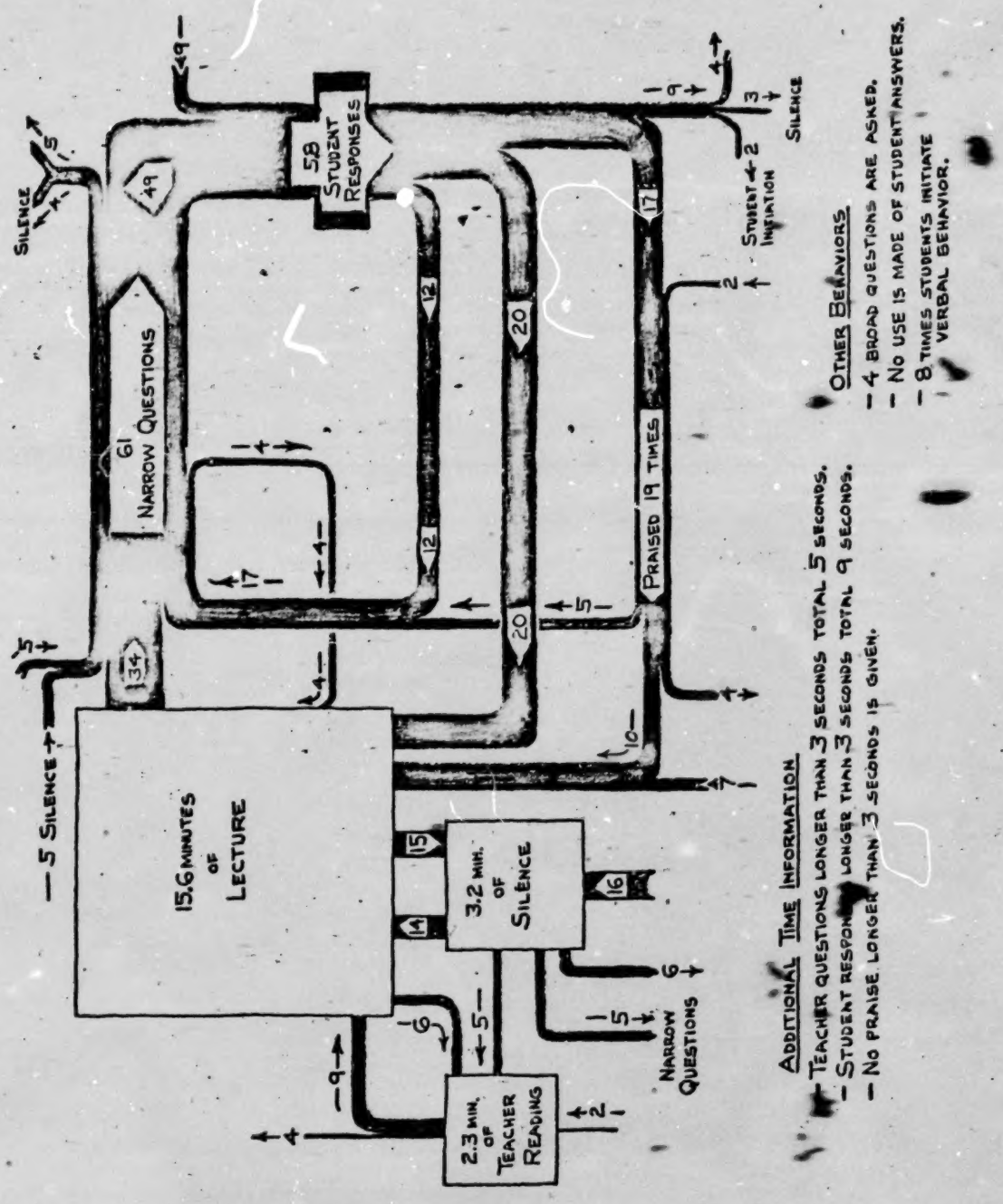
The flow chart in Figure 1 represents another way in which classroom behaviors can be described, based upon observations made with category systems like the FIAC. This type description can also be used as a basis for making cross-national comparisons.

The Meaning of Measurements Resulting from the Use of Category Systems

The question remains as to whether the statistics which result from the use of category systems and subsequent data analysis do actually represent what occurred in the

Figure 1

Flow Chart Illustrating 30 Minutes of FIAC Behaviors:
Based Upon Observations Made in
Fifth Grade Science Classes of 22 Nepalese Schools


15.6 MINUTES OF LECTURE
3.2 MIN. OF SILENCE
2.3 MIN. OF TEACHER READING
61 NARROW QUESTIONS
58 STUDENT RESPONSES
PRAISED 19 TIMES
NARROW QUESTIONS
STUDENT INITIATION
SILENCE
OTHER BEHAVIORS
- 4 BROAD QUESTIONS ARE ASKED.
- NO USE IS MADE OF STUDENT ANSWERS.
- 8 TIMES STUDENTS INITIATE VERBAL BEHAVIOR.
ADDITIONAL TIME INFORMATION
- TEACHER QUESTIONS LONGER THAN 3 SECONDS TOTAL 5 SECONDS.
- STUDENT RESPONSES LONGER THAN 3 SECONDS TOTAL 9 SECONDS.
- NO PRAISE LONGER THAN 3 SECONDS IS GIVEN.

classes observed. Exactly what does "72.4% lecture in Nepal" and "27.2% lecture in the U.S.A." mean? Do these statistics indeed mean that more than twice as much lecture, as defined by Flanders, occurred in Nepal than in the U.S.A.? Can similar inferences be drawn, equally valid for both groups observed, based upon the measurements and analyses made?

The answer appears to be "yes." By using category systems, the extent occurrence of many behaviors can be measured and validly compared from one country to another, from one culture to another.

This does not mean that <u>all</u> category systems are suitable for making comparisons between <u>all</u> cultures, for the concepts involved in some category systems may be too culture bound and not meaningfully comparable from one context to another. It does not mean that measurements made with "non-culture bound" instruments provide valid comparisons in <u>all</u> cases -- for the validity of comparisons is dependent not only upon the instrument used, but on factors such as observer effects upon classroom behaviors, the sampling strategies used, how the instruments are used, observer honestly, distortions related to the time units employed, mistakes made during data analysis, and whether certain indices derived from the basic recordings are equally applicable for drawing inferences about all situations observed.

However, category systems <u>do</u> seem to provide a potential for making valid and precise cross-national comparisons not equaled by other techniques used to measure and describe classroom behaviors.

Logically, it seems that category systems can provide this potential, for they make use of standard magnitudes of

physical time and the explicit language of counting, as compared to psychological languages and magnitudes such as "seldom" or "strong emphasis" which may vary from culture to culture and from time to time. Przeworski and Teune (1970) have pointed out:

> Whether two or more phenomena are "comparable" depends on whether their properties have been expressed in a standard language. A language of measurement defines classes of phenomena by providing specific criteria for deciding whether an observation can be assigned to a particular class.... It is a standard language if it can be consistently applied to all individuals or social units.... Classifying observations into categories, ranking them, or counting instances serves to express observations in a language of measurement.... If these observations are expressed in a standard language, they are indeed comparable. (p.93)

Such reasoning supports the choice of category systems for making comparisons. Category systems usually do provide specific, low-inference criteria for deciding whether certain classifications of behavior have occurred or not. Based (usually) upon the frequency of classification, the extent occurrence of these behaviors can be stated using a standard language. Therefore, category systems do appear suitable for making comparisons.

The logical suitability of category systems is also supported by my own use of these instruments within the U.S.A. and Nepal. That is, in these different cultural contexts, the category systems used appeared to measure the same types of behavior, and many (but not all) analyses made did appear to result in descriptions and inferences equally applicable to both countries studied. In addition, the category systems appeared to be used the same way in both countries, since Nepalese observers reached high levels of agreement with

my own observations, and my usage in turn agreed highly with U.S. standards of usage.

It is for these reasons, based upon actual use and logical examination, that category systems seem suitable for use by researchers wishing to make valid cross-national comparisons. It is for these reasons that category systems seem to deserve our consideration in order to determine (1) whether increased efforts should be made to learn the strengths and limitations of these instruments when used to compare behaviors across cultures and nations, and (2) whether their apparent suitability means that comparative educators should increasingly use these instruments when making cross-national comparisons of classroom behaviors.[6]

Notes

[1]Brislin, Lonner, and Thorndike's book (1973), for example, contains chapters entitled "Questionnaire wording and translation," and the "Cross-cultural use of psychological tests;" Warwick and Osherson's book (1973) similarly contains chapters entitled "The random probe: A technique for evaluating the validity of closed questions," "On the comparability of meaningful stimuli in cross-cultural research," and "Asking questions cross-culturally: Some problems of linguistic comparability."

[2]This is not to imply that procedures which result in such descriptions, particularly "participant observation" as practiced by anthropologists, do not have a definite place when classroom observation research is conducted. Indeed, as Pelto (1970) has pointed out:

> The relatively unsystematized scanning of information through participant observation is basic to all the other, more refined, research techniques. The preliminary data from participant observation provide the fieldworker with insights and clues necessary for developing...specialized tools. Participant observation also provides the further checking and monitoring of field information that is necessary for evaluating data gathered by the specialized techniques. (pp.91-92)

In addition to providing information needed to develop, modify, and judge the suitability of specialized tools such as category systems, participant observation also serves as a basis for generating hypotheses about previously unresearched settings and conditions (Dunkin & Biddle, 1974, pp.58-59). The technique is not, however, very suitable for providing data upon which precise cross-national comparisons of classroom behaviors can be made.

[3]Another type of instrument called "sign systems"

may also be distinguished. Category systems and sign systems are similar in many ways, although "no general term is available to distinguish these two types from rating systems" (Rosenshine & Furst, 1973, p.133). Thus, much of what is said about one may apply to the other. Sign systems, however, have been used much less than category systems, are less precise instruments, and do not lend themselves as well to the study of sequential events in the classroom (Dunkin & Biddle, 1974, pp.71-72).

[4]In addition to halo effects, which frequently affect ratings, Kerlinger (1964) points out other types of errors often associated with rating scales. These include the error of severity, "a general tendency to rate all individuals too low on all characteristics"; the error of leniency, an "opposite tendency to rate too high"; and the error of central tendency, a "general tendency to avoid all extreme judgments and rate right down the middle of a rating scale" (p.517).

[5]The category systems described are but two of hundreds that have been developed in recent years to measure aspects of classroom behavior such as the logic and linguistics of classroom occurrences, the classroom as a social system, and techniques of management and control. Readers wishing to learn more about other category systems now existent can profitably look at Dunkin and Biddle (1974) or Simon and Boyer (1970).

[6]Readers may be interested in knowing that a number of descriptive studies using category systems have been conducted in different cultural settings. The FIAC, for example, has been used by Payne (1970) to study classroom behaviors in Amish and non-Amish schools; by Peek (1970) to compare verbal behaviors occurring in a predominantly

Negro high school with those occurring in a predominantly white high school; by Kim (1972) to describe teaching in Korean elementary schools; by Ledesma (1973) to study the verbal behaviors of Filipino elementary teachers and students; and by Pfau, Mathema, Pradhan, Rayamajhi, and Shrestha (in press) to study classroom behaviors occurring in Nepalese elementary and secondary schools.

References

Adams, R. W. Perceived teaching styles. Comparative Education Review, 1970, 14(1), 50-59.

Brislin, R. W., Lonner, W. J., & Thorndike, R. M. Cross-cultural research methods. New York: John Wiley, 1973.

Caldwell, H. E. Evaluation of an in-service methods course by systematic observation of classroom activities. Project No. 6-8760, U. S. Department of Health, Education, & Welfare, Office of Education, 1967. (ERIC Document Reproduction Service No. ED 024 615)

Caldwell, H. E. Evaluation of an in-service science methods course by systematic observation of classroom activities (Doctoral dissertation, Syracuse University, 1968). Dissertation Abstracts, 1969, 29, 4346A-4347A. (University Microfilms No. 69-7725)

Comber, L. C., & Keeves, J. P. Science education in nineteen countries: An empirical study. New York: Halsted Press, 1973.

Deutscher, I. Asking questions cross-culturally: Some problems of linguistic comparability. In D. P. Warwick & S. Osherson (Eds.), Comparative research methods. Englewood Cliffs: Prentice-Hall, 1973.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

Flanders, N. A. et al. Teacher influence patterns and pupil achievement in the second, fourth, and sixth grade levels

(2 vols.). Ann Arbor: University of Michigan, December 1969.

Flanders, N. A. Analyzing teaching behavior. Reading, Mass.: Addison-Wesley, 1970.

Grimshaw, A. D. Comparative sociology: In what ways different from other sociologies? In A. Armer & A. D. Grimshaw (Eds.), Comparative social research: Methodological problems and strategies. New York: John Wiley, 1973.

Husen, T. (Ed.), International study of achievement in mathematics (2 vols.). New York: John Wiley, 1967.

Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart, & Winston, 1964.

Kim, C. S. A preliminary normative study on teaching behavior in elementary school. Seoul: College of Education, Seoul National University, 1972.

Ledesma, C. P. The use of interaction analysis in assessing the verbal behavior of Filipino elementary teachers and pupils (Doctoral dissertation, Boston University, 1972). Dissertation Abstracts International, 1973, 34, 1771A. (University Microfilms No. 73-23,580)

Medley, D. M., & Mitzel, H. E. Measuring classroom behaviors by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Medley, D. M. Some notes on the future history of research in teaching. Classroom Interaction Newsletter, 1967, 2(2), 1-3.

Payne, J. I. Analysis of teacher-student classroom interaction in Amish and non-Amish schools using Flanders interaction analysis technique (Doctoral dissertation, Pennsylvania State University, 1970). Dissertation Abstracts International, 1971, 32, 251A. (University Microfilms No. 71-16,650)

Peek, D. A. A comparison of the verbal behaviors of teachers and pupils in a predominantly Negro high school with the verbal behaviors of teachers and pupils in a predominantly white high school (Doctoral dissertation, East Texas State University, 1970). Dissertation Abstracts International, 1971, 31, 4630A. (University Microfilms No. 71-8648)

Pelto, P. J. Anthropological research: The structure of inquiry, New York: Harper & Row, 1970.

Pfau, R. H., Mathema, S. S. B., Pradhan, K. M., Rayamajhi, L. B., & Shrestha, S. K. A national study of classroom behaviors in Nepalese schools: Introduction, methodology, summary of findings, and general recommendations (Vol. 1). Final Report. Kirtipur, Nepal: Research Division, Institute of Education (in press)

Przeworski, A., & Teune, H. The logic of comparative social inquiry. New York: Wiley-Interscience, 1970.

Reed, H. B., & Reed, M. J. Nepal in transition: Educational innovations. Pittsburgh: University of Pittsburgh Press, 1968.

Remmers, H. H. Rating methods in research on teaching. In N. L. Gage (Ed.). Handbook of research on teaching. Chicago: Rand McNally, 1963.

Rosenshine, B. Evaluation of classroom instruction. Review of Educational Research, 1970, 40, 279-300.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In M. W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Simon, A., & Boyer, E. G. (Eds.), Mirrors for behavior II: An anthology of observational instruments. Classroom Interaction Newsletter (2 vols.), Special Edition, 1970.

Warwick, D. P., & Osherson, S. (Eds.), Comparative research methods. Englewood Cliffs: Prentice-Hall, 1973.

Appendix A

Explanations of FIAC Calculations

Category Percents (Categories 1 to 10). Each calculation represents the occurrence of a particular category of behavior, expressed as a percent of the total category numbers recorded.

Teacher Talk. This calculation indicates the total percentage of teacher talk which occurred during the periods of classroom interaction recorded.

Student Talk. This calculation indicates the total percentage of student talk which occurred during the periods of classroom interaction recorded.

S/(S+T) Ratio. This ratio is a measure of the relative amount of student talk to teacher talk. It will be higher in classes where students do more talking. A ratio of 0.5 means that the students and teachers talked for equal amounts of time; a ratio of 0.1 means that one-tenth of the talking was done by students.

I/(I+D) Ratio. This ratio indicates the relative amount of indirect teacher talk (Categories 1+2+3+4) that occurred to the amount of direct teacher talk (Categories 5+6+7). This ratio will be higher when relatively more indirect teacher behaviors occur. A ratio of 0.33 means that for every one indirect category number recorded, two direct category numbers were recorded.

i/(i+d) Ratio. This ratio eliminates the effects of Categories 4 and 5 (asks questions; lecturing), but is

otherwise similar to the I/(I+D) Ratio. The i/(i+d) ratio has been defined as an index which corresponds to the teachers' tendency to react to the ideas and feelings of students (Flanders, 1970. p.102). The more use made of student ideas, and the more praise and encouragement given by the teacher, compared to giving directions and criticism, the greater will be this ratio.

Extended Indirect Influence. This calculation indicates teacher tendencies to praise, encourage, and accept or use the ideas of students for longer than three seconds at a time, and to move from one of these types of behavior to another. The greater this calculation, the more such behaviors occurred

Extended Direct Influence. This calculation indicates teacher tendencies to criticize and give directions for longer than three seconds, and to move from one of these types of behavior to the other. The greater this calculation, the more such behaviors occurred.

Steady State Ratio (10x10). The steady state ratio reflects the tendency of teacher and student talk to remain the same for periods lasting longer than three seconds. The higher this ratio is, the less rapid were changes from one FIAC behavior to another.

Pupil Steady State Ratio. The pupil steady state ratio is an index reflecting the rapidity of the teacher-student interchange. The higer this ratio, the more students expressed themselves in a sustained manner.

Cell 9-9. This calculation indicates the degree to which students expressed their own ideas and talk they initiated in a sustained manner.

Pupil Initiation Ratio. This calculation indicates what percentage of student talk was judged by an observer to be an act of initiation.

Content Cross Ratio. A high content cross ratio is an indication that the main focus of class discussion was on subject matter, that the teacher took a very active roll in the discussion, and that attention to motivation and discipline problems was at a minimum (Flanders, 1970, p.106).

Drill. This calculation indicates the percent of interaction recorded during which (a) student talk occurred in direct response to teacher questions, and (b) such student talk was followed by a question based on teacher ideas.

Cell 3-3. This calculation indicates the degree to which teachers accepted or used student ideas in a sustained manner.

Teacher Question Ratio (TQR). This ratio is defined as "an index representing the tendency of a teacher to use questions when guiding the more content oriented part of the class discussion" (Flanders, 1970, p.102).

Instantaneous Teacher Question Ratio (TQR89). This ratio indicates the tendency of teachers to respond to student talk with questions based on teacher ideas, compared to a tendency to lecture.

Instantaneous Teacher Response Ratio (TRR89). The TRR89 indicates the tendency of teachers to respond to student talk with praise or encouragement or by integrating student talk into the class discussion, compared to teacher tendencies to criticize or give directions immediately after students stop speaking.